# Small diameters and generators for arithmetic lattices in $\mathrm{SL}_2(\mathbb{R})$ and certain Ramanujan graphs

**Raphael S. Steiner[1]**

## Abstract

We show that arithmetic lattices in $\mathrm{SL}_2(\mathbb{R})$, stemming from the proper units of an Eichler order in an indefinite quaternion algebra over $\mathbb{Q}$, admit a 'small' covering set. In particular, we give bounds on the diameter if the quotient space is co-compact. Consequently, we show that these lattices admit small generators. Our techniques also apply to definite quaternion algebras where we show Ramanujan-strength bounds on the diameter of certain Ramanujan graphs without the use of the Ramanujan bound.

## 1 Introduction

Let $B$ be an indefinite quaternion algebra over $\mathbb{Q}$ of reduced discriminant[1] $D$ and $R \subset B$ be an Eichler order of level $Q$, that is an order $R$ such that locally for each prime $p \nmid D$ $R_p \cong \begin{pmatrix} \mathbb{Z}_p & \mathbb{Z}_p \\ Q\mathbb{Z}_p & \mathbb{Z}_p \end{pmatrix}$ and for $p \mid D$ $R_p$ is the unique maximal order in $B_p$. Let $\Gamma$ be the subset of proper units (elements of norm one) of $R$. Further, fix an isomorphism $B(\mathbb{R})$ with the matrix algebra $\mathrm{Mat}_{2\times 2}(\mathbb{R})$.[2] In the case where $B$ is already split over $\mathbb{Q}$ ($D = 1$), we may choose this identification such that

---

[1] Product of the finite primes at which $B$ ramifies.
[2] Any other isomorphism is a conjugate by a matrix in $\mathrm{GL}_2(\mathbb{R})$.

✉ Raphael S. Steiner
  raphael.steiner.academic@gmail.com

1  Department of Mathematics, ETH Zürich, 8092 Zürich, Switzerland

$$\Gamma = \Gamma_0(Q) = \left\{ \left( \begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right) \in \underset{2}{\mathrm{SL}}(\mathbb{Z}) \,|\, c \equiv 0 \mod (Q) \right\}$$

is the familiar congruence lattice. Under the identification of $B(\mathbb{R})$, $\Gamma$ is a lattice in $\mathrm{SL}_2(\mathbb{R})$ of co-volume $V_\Gamma = (DQ)^{1+o(1)}$. Furthermore, $\Gamma$ is co-compact if and only if $B(\mathbb{Q})$ is a division algebra ($D \neq 1$). We refer to [4] for further details on the above.

A typical example of a fundamental domain for the action of $\Gamma$ on the homogeneous space $\mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2(\mathbb{R}) \cong \mathbb{H}$, the upper half-plane, is a normal polygon, also known as a Dirichlet domain, which is given as follows. Let $w \in \mathbb{H}$ be a point whose stabiliser $\Gamma_w$ in $\Gamma$ consists only of plus and minus the identity element. Then, the normal polygon with centre $w$ is given by

$$\mathcal{F}_{\Gamma,w} = \{z \in \mathbb{H} \,|\, d(z, w) < d(\gamma z, w), \ \forall \gamma \in \Gamma, \gamma \neq \pm I\}, \tag{1.1}$$

where $d$ denotes the hyperbolic distance. If $B$ is split and $\Gamma = \Gamma_0(Q)$, then another typical example of a fundamental domain is given by the standard polygon:

$$\mathcal{F}_{\Gamma,\infty} = \left\{ z \in \mathbb{H} \,|\, |\Re(z)| < \tfrac{1}{2}, \ \Im(z) > \Im(\gamma z), \ \forall \gamma \in \Gamma - \Gamma_\infty \right\}, \tag{1.2}$$

also referred to as a Ford domain. The sides of these polygons may be paired up such that the corresponding side-pairing motions together with $-I$ generate the group $\Gamma$ [12, Chap. 2]. Thus, it comes as no surprise that the 'size' of these fundamental domains, e.g. the diameter of $\Gamma \backslash \mathbb{H}$ if the latter is compact, are related to the size of generators of $\Gamma$. Algorithms to compute fundamental domains and, subsequently, a set of generators have been devised by Johansson [14], Voight [32], and subsequently improved by Rickards [26] if $\Gamma$ is co-compact, and Kurth–Long [18] if $\Gamma$ is a finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$ through the use of Farey symbols [17]. In the latter case, further algorithms based on the Reidemeister–Schreier process [25, 30] are available to determine an independent set of generators for $\Gamma(p)$, where $p$ is a prime, by Frasch [8], for $\Gamma_0(p)$ by Rademacher [23], and for $\Gamma_0(Q)$, for general $Q \in \mathbb{N}$, by Chuman [3]. Albeit the former algorithms due Johansson, Voight, Rickards, and Kurth–Long work well in practise, they don't give any answer to the question regarding the asymptotic size of the (produced) generators and, thus, an upper bound on their time complexity. The algorithms based on the Reidemeister–Schreier process do give an explicit set of generators whose elements are of polynomial size in the co-volume, but they are far from the generating set whose elements are of least size. Stronger results in that direction were given by Khoai [16], who showed that $\Gamma_0(Q)$ is generated by elements of Frobenius norm bounded by $O(Q^2)$, respectively $O(Q)$ if $Q$ is a prime power, and Chu–Li [2] who managed to show that $\Gamma$ co-compact is generated by its elements of Frobenius norm bounded by $O_\epsilon(V_\Gamma^{2.56+\epsilon})$.[3] In this paper, we shall prove the following theorems.

**Theorem 1** *$\Gamma_0(Q)$ is generated by its elements of Frobenius norm $O_\epsilon(Q^{1+\epsilon})$.*

---

[3] They state their theorem with exponent 7.68, though their method gives $5.12 + o(1)$. In turn, this can be halved again by replacing their final argument by the argument in this paper.

**Theorem 2** *Let $\Gamma \subset SL_2(\mathbb{R})$ be a co-compact arithmetic lattice of co-volume $V_\Gamma$ stemming from the proper units of an Eichler order $R$ of level $Q$ in a quaternion algebra $B$ over $\mathbb{Q}$ of reduced discriminant $D$. Then, for almost every $\sigma \in \Gamma \backslash SL_2(\mathbb{R})$, $\sigma^{-1}\Gamma\sigma$ is generated by its elements of Frobenius norm $O_\epsilon(V_\Gamma^{2+\epsilon})$. In other words, almost every embedding $\Gamma$ of the proper units of the order $R$ into $SL_2(\mathbb{R})$ is generated by its elements of Frobenius norm $O_\epsilon(V_\Gamma^{2+\epsilon})$. If one assumes either of the following conditions:*

- *$Q$ is square-free,*
- *Selberg's eigenvalue conjecture for $\Gamma_0(DQ)$,*
- *the sup-norm conjecture in the level aspect for exceptional eigenforms on $\Gamma \backslash SL_2(\mathbb{R})/SO_2(\mathbb{R})$,*

*then, $\Gamma$ is in fact generated by its elements of Frobenius norm $O_\epsilon(V_\Gamma^{2+\epsilon})$ regardless of the embedding.*

Theorem 1 follows from carefully bounding the standard polygon (1.2) by isometric circles of large radius. This is layed out in Sect. 3. Theorem 2 follows from showing that the normal polygon (1.1) is contained in a ball of small radius.

**Theorem 3** *Let $\Gamma \subset SL_2(\mathbb{R})$ be a co-compact arithmetic lattice of co-volume $V_\Gamma$ stemming from the proper units of an Eichler order $R$ of level $Q$ in a quaternion algebra $B$ over $\mathbb{Q}$ of reduced discriminant $D$. Then, for every $\eta > 0$, $w \in \Gamma \backslash \mathbb{H}$ satisfies*

$$\sup_{z \in \mathbb{H}} \min_{\gamma \in \Gamma} d(\gamma z, w) \leq (2 + \eta) \log 3V_\Gamma \tag{1.3}$$

*with probability $1 - o(1)$ as $V_\Gamma \to \infty$. Suppose either of the following statements is true:*

- *$Q$ is square-free,*
- *Selberg's eigenvalue conjecture for $\Gamma_0(DQ)$, or*
- *the sup-norm conjecture in the level aspect for exceptional eigenforms on $\Gamma \backslash SL_2(\mathbb{R}) / SO_2(\mathbb{R})$.*

*Then, for every $\eta > 0$ and $w \in \mathbb{H}$, $z \in \Gamma \backslash \mathbb{H}$ satisfies*

$$\min_{\gamma \in \Gamma} d(\gamma z, w) \leq (1 + \eta) \log 3V_\Gamma \tag{1.4}$$

*with probability $1 - o(1)$ as $V_\Gamma \to \infty$. In particular, the diameter of the hyperbolic surface $\Gamma \backslash \mathbb{H}$ is bounded by $(2 + o(1)) \log 3V_\Gamma$ as $V_\Gamma \to \infty$.*

The bound (1.4) for the almost diameter is sharp and one may speculate whether the actual diameter is around the same length. The latter would imply that the co-compact lattices under consideration are generated by its elements of norm $O_\epsilon(V_\Gamma^{1+\epsilon})$ and thus would bring it onto equal footing with Theorem 1. We should further remark that the bound depends on some Siegel-zero estimates and is, thus, not effective, respectively, can be made effective with at most one exception.

In order to prove Theorem 2, one could make use of Ratner's exponential mixing for $SL_2(\mathbb{R})$ [24]. This approach was taken by Chu–Li [2], who showed (1.3) for every

$w \in \mathbb{H}$ with 2 replaced by 2.56. Instead, we simplify the argument a bit by using the operator which averages over a sphere of a given radius. The latter operator was employed by Golubev–Kamber [10] who showed, amongst many results of related nature, that the almost diameter under the Selberg eigenvalue conjecture is bounded by $\log(3V_\Gamma) + (2 + o(1)) \log \log(9V_\Gamma)$ under the mild assumption that $\Gamma \backslash \mathbb{H}$ has not too many points of small injectivity radius.

Our improvement compared to previous results comes from the incorporation of a density estimate for exceptional eigenvalues as well as the newly available fourth moment bound for Maass forms in the level aspect by Khayutin–Nelson–Steiner [15].

At last, we would also like to touch on the closely related problem of bounding the diameter of expander graphs. A rich family of expander graphs, so-called Ramanujan graphs, may be constructed from arithmetic data associated to *definite* quaternion algebras [20, 22]. These Ramanujan graphs (by definition) enjoy a large spectral gap, which in turn yields a small upper bound on the diameter. The incorporation of a density estimate for large eigenvalues for (homogeneous) expander graphs has proved valuable in showing that they admit a smaller diameter than what could be directly inferred from their spectral gap [9]. In Sect. 4, we demonstrate the usefulness of the fourth moment bound of Khayutin–Nelson–Steiner [15] also in the context of expander graphs, by proving upper bounds on the diameter of certain Ramanujan graphs *without* the use of the Ramanujan bound which are of *equal* strength.

## 2 Co-compact lattices

Write $G = \mathrm{SL}_2(\mathbb{R})$ and $K = \mathrm{SO}_2(\mathbb{R})$ for short. Let $\mu$ denote the Haar measure on $G$ normalised such that $\mathrm{d}\mu(n(x)a(y)k(\theta)) = \frac{1}{y^2}\mathrm{d}x\mathrm{d}y\mathrm{d}\theta$, where

$$n(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \quad a(y) = \begin{pmatrix} y^{\frac{1}{2}} & 0 \\ 0 & y^{-\frac{1}{2}} \end{pmatrix}, \quad k(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Let $\Gamma$ be a co-compact lattice as in Sect. 1. Denote by $V_\Gamma$ the co-volume of $\Gamma$ with respect to $\mu$. Further, $\mu$ descends to a finite measure $\nu$ on $\Gamma \backslash G$, which we normalise to a probability measure. With $\nu_\star$ we denote the push forward of $\nu$ to $\Gamma \backslash G/K$. Let $\{u_j\}_j$ be an orthonormal basis of Hecke–Maass forms on $L^2(\Gamma \backslash G/K, \nu_\star)$, which we may also regard as an orthonormal basis of the $K$-invariant subspace of $L^2(\Gamma \backslash G, \nu)$. We denote the Laplace eigenvalue of $u_j$ with $-\lambda_j = -(\frac{1}{4} + t_j^2)$, where $t_j \in \mathbb{R} \cup i[0, \frac{1}{2}]$. We let $d$ denote the hyperbolic distance on the upper half-plane $\mathbb{H} \cong G/K$ and $u$ the related quantity

$$u(z, w) = \tfrac{1}{2}(\cosh(d(z,w)) - 1) = \frac{|w - z|^2}{4\,\mathrm{Im}(z)\,\mathrm{Im}(w)}.$$

We note that

$$u(z, w) = \tfrac{1}{4}\,\mathrm{tr}(gg^t) - \tfrac{1}{2} =: u(g)$$

where $g$ is any matrix that takes $z$ to $w$ and $u : G \to \mathbb{R}_0^+$ is left and right $K$-invariant. Let $S : \mathbb{R}_0^+ \to [0, 1]$ be a smooth bump function supported on $[0, \delta]$ for some small $\delta > 0$, such that its Selberg/Harish-Chandra transform:

$$h(t) = 4\pi \int_0^\infty S(u) \cdot {}_2F_1(\tfrac{1}{2} + it, \tfrac{1}{2} - it; 1; -u) du$$

is non-negative and

$$h(\pm \tfrac{i}{2}) = 4\pi \int_0^\infty S(u) du \asymp 1, \tag{2.1}$$

where the implied constants may depend on $\delta$. We note that

$$|h(t)| \leq 4\pi \int_0^\infty |S(u)| du = h(\pm \tfrac{i}{2}) \ll 1. \tag{2.2}$$

Let

$$B(g_1, g_2) = \sum_{\gamma \in \Gamma} S(u(g_2^{-1} \gamma g_1)),$$

where $S(u(g_2^{-1} g_1))$ is a point-pair invariant, and thus we get the spectral expansion (cf. [12, Theorem 1.14])

$$B(g_1, g_2) = \frac{2\pi}{V_\Gamma} \sum_j h(t_j) u_j(g_1) \overline{u_j(g_2)}. \tag{2.3}$$

$B_g := B(g, \cdot)$ will take the role of a smooth right $K$-invariant ball on $\Gamma \backslash G$. The following lemma is of crucial importance. Essentially, it says that $B_g$ may be compared to an Euclidean ball.

**Lemma 4** *Suppose $\delta$ is sufficiently small but (strictly) positive. Then, we have for any $g_1 \in G$ that*

$$\int_{\Gamma \backslash G} |B(g_1, g_2)|^2 d\nu(g_2) \ll \frac{1}{V_\Gamma},$$

*where the implied constant may depend on $\delta$.*

**Proof** We have

$$\int_{\Gamma \backslash G} |B(g_1, g_2)|^2 d\nu(g_2) = \frac{4\pi^2}{V_\Gamma^2} \sum_j |h(t_j)|^2 |u_j(g_1)|^2$$

$$= \frac{2\pi}{V_\Gamma} \sum_{\gamma \in \Gamma} ((S \circ u) \star (S \circ u))(g_1^{-1} \gamma g_1),$$

where the convolution is taken on $G$. We note that $(S \circ u) \star (S \circ u)$ is bounded for fixed $\delta$. Moreover, it is supported on $g \in G$ with $u(g) \leq 4\delta(1 + \delta)$. We now note that $u(g) \geq \frac{(\operatorname{tr} g)^2}{4} - 1$. Thus, if $\delta$ is sufficiently small the sum over all hyperbolic $\gamma \in \Gamma$ is zero since for those $\operatorname{tr} \gamma$ is at least 3. Now again for $\delta$ sufficiently small, by the Margulis' Lemma (cf. [7, Sect. 3.1]), the subgroup $\widetilde{\Gamma}$ generated by the remaining $\gamma$ for which $u(g_1^{-1} \gamma g_1) \leq 4\delta(1 + \delta)$ is virtually abelian and in particular of one of the following types:

(i) An infinite cyclic group generated by an hyperbolic or parabolic isometry;
(ii) A finite cyclic group generated by an elliptic isometry;
(iii) An infinite dihedral group generated by two elliptic isometries of order 2.

The first case only contains the elliptic elements plus minus the identity. In the second case, $\widetilde{\Gamma}$ is finite and of order bounded by 6 as the characteristic polynomial is a cyclotomic polynomial of degree at most two over $\mathbb{Q}$. In the third case, the subgroup $\widetilde{\Gamma}$ fixes a geodesic and all elliptic elements not equal to plus minus the identity fix a single point on this geodesic (and perform a rotation by $\pi$ around the point). These fixpoints are evenly spread along the geodesic by the distance corresponding to the translation of the minimal hyperbolic element in $\widetilde{\Gamma}$. Thus, if $\delta$ is small enough, there are at most four elliptic elements $\gamma \in \widetilde{\Gamma}$ such $u(g_1^{-1} \gamma g_1) \leq 4\delta(1 + \delta)$. We conclude the lemma. $\qquad\square$

We now get to the heart of the argument. We shall use an averaging operator which averages over a sphere of radius $T$:

$$
A_T f(z) = \frac{1}{2\pi} \int_0^{2\pi} f\left(g_z k(\theta) \begin{pmatrix} e^{\frac{T}{2}} & 0 \\ 0 & e^{-\frac{T}{2}} \end{pmatrix}\right) d\theta,
$$

where $g_z$ is any matrix that takes $i$ to $z$.

We shall prove the following proposition.

**Proposition 5** *For $T \geq 1$, we have*

$$
\int_{\Gamma \backslash G} \sup_{g_2 \in \Gamma \backslash G / K} \left| \langle A_T B_{g_1}, B_{g_2} \rangle - \langle B_{g_1}, u_0 \rangle \overline{\langle B_{g_2}, u_0 \rangle} \right|^2 d\nu(g_1)
$$
$$
\ll_\epsilon T^2 V_\Gamma^{-2+\epsilon} \left( e^{-T} + e^{-\frac{T}{2}} V_\Gamma^{-1} \right). \tag{2.4}
$$

*Assume either of the following conditions:*

- *Q is square-free,*
- *Selberg's eigenvalue conjecture for $\Gamma_0(DQ)$,*
- *the sup-norm conjecture in the level aspect for exceptional eigenforms on $\Gamma \backslash SL_2(\mathbb{R}) / SO_2(\mathbb{R})$,*

*then we have the stronger bound*

$$
\| A_T B_g - \langle B_g, u_0 \rangle u_0 \|_2^2 \ll_\epsilon T^2 V_\Gamma^{-2+\epsilon} \left( e^{-T} V_\Gamma + e^{-\frac{T}{2}} V_\Gamma^{\frac{1}{2}} \right). \tag{2.5}
$$

Before proceeding with the proof of the proposition, we shall show how Theorem 3 follows from it. The first half of the Theorem follows from choosing $T = (2 + o(1)) \log 3 V_\Gamma$ in (2.4). As the main term

$$\langle B_{g_1}, u_0 \rangle \overline{\langle B_{g_2}, u_0 \rangle} = \frac{4\pi^2}{V_\Gamma^2} |h(\tfrac{i}{2})|^2 \asymp V_\Gamma^{-2},$$

it follows that $\langle A_T B_{g_1}, B_{g_2} \rangle > 0$ for most $g_1$. In order to prove the second half of the theorem, we shall use (2.5) with $T_0 = (1 + o(1)) \log 3 V_\Gamma$. We then find

$$\nu_\star(\{z \in \Gamma \backslash G/K \mid A_{T_0} B_g(z) = 0\}) \ll V_\Gamma^2 \|A_{T_0} B_g - \langle B_g, u_0 \rangle u_0 \|_2^2 = o(1).$$

**Proof of Proposition 5** We first note that each $u_i$ is an eigenfunction of $A_T$ with eigenvalue $_2F_1(\tfrac{1}{2} - it_i, \tfrac{1}{2} + it_i; 1; \tfrac{1}{2} - \tfrac{1}{2} \cosh(T)) = P_{-\frac{1}{2}-it_i}(\cosh(T))$, cf. [12, Corollary 1.13]. We have that this eigenvalue is bounded by

$$\begin{cases} (T+1)e^{-\frac{T}{2}(1-\sqrt{1-4\lambda_i})}, & 0 \le \lambda_i \le \tfrac{1}{4}, \\ (T+1)e^{-\frac{T}{2}}, & \tfrac{1}{4} \le \lambda_i, \end{cases}$$

cf. [10, Props. 2.3 & 7.2]. Hence, after referring to the spectral expansion (2.3), we find that

$$\langle A_T B_{g_1}, B_{g_2} \rangle - \langle B_{g_1}, u_0 \rangle \overline{\langle B_{g_2}, u_0 \rangle}$$

$$= O\left( \frac{1}{V_\Gamma^2} \sum_{0 < \lambda_j < \frac{1}{4}} T \left(e^{-\frac{T}{2}}\right)^{1-\sqrt{1-4\lambda_j}} |h(t_j)|^2 |u_j(g_1) u_j(g_2)| \right)$$

$$+ O\left( T e^{-\frac{T}{2}} \|B_{g_1}\|_2 \|B_{g_2}\|_2 \right) \tag{2.6}$$

for any two $g_1, g_2 \in G$ and $T \ge 1$. We note that the second smallest eigenvalue $\lambda_1 \ge \frac{3}{16}$, due to Selberg [31] and the Jacquet–Langlands correspondence [13]. By Lemma 4, we find that the second error term is bounded by $O(T e^{-\frac{T}{2}} V_\Gamma^{-1})$. It remains to deal with the first error term. By Cauchy–Schwarz, we may estimate

$$\frac{1}{V_\Gamma^2} \sum_{0 < \lambda_j < \frac{1}{4}} T \left(e^{-\frac{T}{2}}\right)^{1-\sqrt{1-4\lambda_j}} |h(t_j)|^2 |u_j(g_1) u_j(g_2)|$$

$$\le \left( \frac{1}{V_\Gamma^2} \sum_{\frac{3}{16} \le \lambda_j < \frac{1}{4}} |h(t_j)|^2 |u_j(g_2)|^2 \right)^{\frac{1}{2}}$$

$$\times \left( \frac{1}{V_\Gamma^2} \sum_{\frac{3}{16} \le \lambda_j < \frac{1}{4}} T^2 \left(e^{-\frac{T}{2}}\right)^{2(1-\sqrt{1-4\lambda_j})} |h(t_j)|^2 |u_j(g_1)|^2 \right)^{\frac{1}{2}}$$

$$\ll \|B_{g_2}\|_2 \left( \frac{1}{V_\Gamma^2} \sum_{\frac{3}{16} \le \lambda_j < \frac{1}{4}} T^2 \left( e^{-\frac{T}{2}} \right)^{2(1-\sqrt{1-4\lambda_j})} |h(t_j)|^2 |u_j(g_1)|^2 \right)^{\frac{1}{2}}$$

$$\ll \frac{1}{V_\Gamma^{\frac{1}{2}}} \left( \frac{1}{V_\Gamma^2} \sum_{\frac{3}{16} \le \lambda_j < \frac{1}{4}} T^2 \left( e^{-\frac{T}{2}} \right)^{2(1-\sqrt{1-4\lambda_j})} |h(t_j)|^2 |u_j(g_1)|^2 \right)^{\frac{1}{2}}.$$

Thus, we find

$$\int_{\Gamma \backslash G} \sup_{g_2 \in \Gamma \backslash G / K} \left| \langle A_T B_{g_1}, B_{g_2} \rangle - \langle B_{g_1}, u_0 \rangle \overline{\langle B_{g_2}, u_0 \rangle} \right|^2 d\nu(g_1)$$

$$\ll \frac{1}{V_\Gamma^3} \left( \sum_{\frac{3}{16} \le \lambda_j < \frac{1}{4}} T^2 \left( e^{-\frac{T}{2}} \right)^{2(1-\sqrt{1-4\lambda_j})} |h(t_j)|^2 \int_{\Gamma \backslash G} |u_j(g_1)|^2 d\nu(g_1) \right)$$

$$+ T^2 e^{-T} V_\Gamma^{-2}$$

$$\ll \frac{1}{V_\Gamma^3} \left( \sum_{\frac{3}{16} \le \lambda_j < \frac{1}{4}} T^2 \left( e^{-\frac{T}{2}} \right)^{2(1-\sqrt{1-4\lambda_j})} \right) + T^2 e^{-T} V_\Gamma^{-2}.$$

The sum over the exceptional eigenvalues we may bound using the density estimate (cf. [12, Theorem 11.7])

$$\sharp\{j > 0 \mid s_j > \sigma\} \ll_\epsilon (DQ)^{3-4\sigma+\epsilon}, \tag{2.7}$$

for any $\sigma \ge \frac{1}{2}$, where $s_j = \frac{1}{2} - it_j$. We note that this density estimate, which a priori holds for $\Gamma_0(DQ)$, also holds for the lattice $\Gamma$ under consideration. This is the case since we may consider an explicit Jacquet–Langlands transfer sending a form on $\Gamma$ to its corresponding newform form on $\Gamma_0(N) \supseteq \Gamma_0(DQ)$ for some $D \div N \div DQ$ which has the same eigenvalue. Under this map, at most $Q^{o(1)}$ forms get mapped onto the same image. This follows from [1, Theorem 1 & its Corollary, Theorem 4] and noting that at the places dividing $D$ the corresponding representations are one-dimensional. Using the estimate (2.7), we find

$$\sum_{\frac{3}{16} \le \lambda_j < \frac{1}{4}} T^2 \left( e^{-\frac{T}{2}} \right)^{2(1-\sqrt{1-4\lambda_j})} = T^2 \sum_{\frac{1}{2} < s_j \le \frac{3}{4}} \left( e^{\frac{T}{2}} \right)^{4(s_j-1)}$$

$$\ll_\epsilon (DQ)^\epsilon T^2 \left( e^{-T}(DQ) + e^{-\frac{T}{2}} \right)$$

$$\ll_\epsilon V_\Gamma^\epsilon T^2 \left( e^{-T} V_\Gamma + e^{-\frac{T}{2}} \right), \tag{2.8}$$

Hence, we conclude the first part of the proposition. For the second part, we need to bound

$$\|A_T B_g - \langle B_g, u_0\rangle u_0\|_2^2$$
$$\ll \frac{1}{V_\Gamma^2} \sum_{\frac{3}{16}\le\lambda_j<\frac{1}{4}} T^2 \left(e^{-\frac{T}{2}}\right)^{2(1-\sqrt{1-4\lambda_j})} |h(t_j)|^2 |u_j(g)|^2 + T^2 e^{-T} \|B_g\|_2^2.$$

We note that the second summand is $\ll T^2 e^{-T} V_\Gamma^{-1}$, which is satisfactory. Furthermore, we find that the first summand is empty if we assume Selberg's eigenvalue conjecture, which takes care of that case. In the other cases, we first recall (2.8), which implies there is a $g_0 \in \Gamma\backslash G$ such that

$$\frac{1}{V_\Gamma^2} \sum_{\frac{3}{16}\le\lambda_j<\frac{1}{4}} T^2 \left(e^{-\frac{T}{2}}\right)^{2(1-\sqrt{1-4\lambda_j})} |h(t_j)|^2 |u_j(g_0)|^2 \ll_\epsilon T^2 V_\Gamma^{-2+\epsilon} \left(e^{-T} V_\Gamma + e^{-\frac{T}{2}}\right).$$

Thus, in order to prove (2.5), it is sufficient to bound

$$\frac{1}{V_\Gamma^2} \sum_{\frac{3}{16}\le\lambda_j<\frac{1}{4}} T^2 \left(e^{-\frac{T}{2}}\right)^{2(1-\sqrt{1-4\lambda_j})} |h(t_j)|^2 \left(|u_j(g)|^2 - |u_j(g_0)|^2\right)$$
$$\le \frac{1}{V_\Gamma^2} \left(\sum_{\frac{3}{16}\le\lambda_j<\frac{1}{4}} T^4 \left(e^{-\frac{T}{2}}\right)^{4(1-\sqrt{1-4\lambda_j})}\right)^{\frac{1}{2}} \left(\sum_{\frac{3}{16}\le\lambda_j<\frac{1}{4}} \left(|u_j(g)|^2 - |u_j(g_0)|^2\right)^2\right)^{\frac{1}{2}}.$$

If we assume that the level $Q$ of $R$ is square free, then [15, Theorem 1.10] shows that

$$\sum_{\frac{3}{16}\le\lambda_j<\frac{1}{4}} \left(|u_j(g)|^2 - |u_j(g_0)|^2\right)^2 \ll_\epsilon V_\Gamma^{1+\epsilon}.$$

The same conclusion also holds if we assume the sup-norm conjecture and referring to Weyl's law. Finally, we may estimate the first factor by once more referring to the density estimate (2.7):

$$\sum_{\frac{3}{16}\le\lambda_j<\frac{1}{4}} T^4 \left(e^{-\frac{T}{2}}\right)^{4(1-\sqrt{1-4\lambda_j})} \ll_\epsilon V_\Gamma^\epsilon T^4 \left(e^{-2T} V_\Gamma + e^{-T}\right).$$

We conclude the proof. □

We are left to infer Theorem 2 from Theorem 3. Suppose the stabiliser of the point $i \in \mathbb{H}$ consists only of plus minus the identity and consider the normal polygon $\mathcal{F}_{\Gamma,i}$. Suppose further that $\mathcal{F}_{\Gamma,i}$ is contained in a ball of radius $r$. We shall now translate this

picture from the upper half-plane to the Poincaré disk through the Cayley transforma-
tion $\phi : \mathbb{H} \to \mathfrak{D}$, which maps $z \mapsto (z - i)/(z + i)$. Under this map, $\mathcal{F}_{\Gamma, i}$ gets mapped
to a Ford domain $\mathfrak{F}_\Gamma$. A motion $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R})$ gets transferred to the motion

$$\gamma^\phi = \phi \gamma \phi^{-1} = \begin{pmatrix} \frac{a+d}{2} + i\frac{b-c}{2} & \frac{a-d}{2} - i\frac{b+c}{2} \\ \frac{a-d}{2} + i\frac{b+c}{2} & \frac{a+d}{2} - i\frac{b-c}{2} \end{pmatrix} =: \begin{pmatrix} \overline{F} & \overline{E} \\ E & F \end{pmatrix} \tag{2.9}$$

in $\mathrm{SU}(1, 1)$. Now, Ford [6] proved that $\Gamma^\phi = \phi \Gamma \phi^{-1}$ is generated by the motions
$\gamma^\phi \in \Gamma^\phi$ whose partial arc of its isometric circle forms part of the boundary of $\mathfrak{F}_\Gamma$.
We have that the isometric circle corresponding to the motion (2.9) is given by the
equation $|Ez + F|$, a circle with radius $1/|E|$ and centre $-F/E$. Thus, in order for
the isometric circle of $\gamma^\phi$ to intersect $\mathfrak{F}_\Gamma$ one must have $(|F| - 1)/|E| \leq \mathrm{artanh}(\frac{r}{2})$,
which after a calculation yields $|E| \leq \sinh(r)$ and thus

$$\|\gamma\|_F^2 = 4|E|^2 + 2 \ll e^{2r}.$$

Thus, we may conclude Theorem 2 from Theorem 3 after noting that the points $z \in \mathbb{H}$
with non-trivial stabiliser are a null-set; hence, they may be excluded for the first part
and for the second part one may conjugate the group by a tiny bit if $i$ happens to be
such a point.

## 3 Non-co-compact lattices

In this section, we shall bound the standard polygon (1.2). We shall also point out that
a similar argument has been carried out in the appendix to [28].

We start with two preparatory Lemmata.

**Lemma 6** *For any $\epsilon > 0$, we may find a constant $C_\epsilon > 0$ with the following property.
For two relatively prime integers $a, b$ and natural number $D$, we may find a natural
number $k \leq C_\epsilon D^\epsilon$ such that $(a + kb, D) = 1$.*

**Proof** This is [27, Lemma 2.1]. $\square$

**Lemma 7** *For any $\epsilon > 0$, we may find a constant $C_\epsilon > 0$ with the following property.
For any real number $x$ and natural number $D$, we either have that*

(i) *there is an integer $c$ such that $|x - c| \leq \frac{1}{2(1 + C_\epsilon D^\epsilon)}$, or that*

(ii) *there is a natural number $b \leq 2(1 + C_\epsilon D^\epsilon)^2$ and an integer $(a, bD) = 1$ such
that*

$$|bx - a| \leq \tfrac{1}{2}.$$

**Proof** By Dirichlet's Approximation Theorem, we may find $d \in \mathbb{N}$ with $d \leq K = 2(1 + C_\epsilon D^\epsilon)$ and $c \in \mathbb{Z}$ such that $|dx - c| \leq \frac{1}{K}$. Without loss of generality, we

may assume $(c, d) = 1$. Now, if $d = 1$, then we are done as the first condition is satisfied. Suppose now that $d \geq 2$, then we may find a pair of integers $a, b$ such that $ad - bc = \pm 1$, where the sign is chosen such that $\frac{a}{b}$ and $x$ lie on the same side of $\frac{c}{d}$ on the number line. We note that the pair of integers $(a + kc, b + kd)$, where $k$ is any integer, also satisfies the same equation. Hence, we may assume $0 < b \leq d$. Note that we have $(a, c) = 1$ and hence we may apply Lemma 6 to even require $(a, bD) = 1$ at the cost of increasing the size of $b$ to at most $(1 + C_\epsilon D^\epsilon)d$. We have

$$|x - \tfrac{a}{b}| \leq \max \left\{ \frac{1}{dK}, \frac{1}{bd} \right\} \leq \frac{1}{2b}.$$

The conclusion follows. $\qquad \square$

We note that the standard polygon (1.2) agrees with the Ford domain

$$\mathcal{F}_{\Gamma, \infty} = \{z \in \mathbb{H} \mid |\Re(z)| \leq \tfrac{1}{2} \text{ and } |cQz + d| \geq 1, \ \forall (cQ, d) = 1\}. \qquad (3.1)$$

We proceed by showing that the isometric circles $|cQz + d| = 1$, corresponding to the motion $\left( \begin{smallmatrix} a & b \\ cQ & d \end{smallmatrix} \right) \in \Gamma$, that form part of the boundary of $\mathcal{F}_{\Gamma, \infty}$ must have their radius $1/|cQ|$ bounded below by $Q^{-1+o(1)}$.

Let $z = x + iy \in \mathcal{F}_{\Gamma, \infty}$. We now apply Lemma 7 to $xQ$ with $D = Q$. Thus, we have either

(i) $|Qx - c| \leq \frac{1}{2(1 + C_\epsilon Q^\epsilon)}$ for some integer $c$, or
(ii) $|bQx - a| \leq \frac{1}{2}$ some natural number $b \leq 2(1 + C_\epsilon Q^\epsilon)^2$ and integer $a$ with $(a, bQ) = 1$.

Let us first deal with the second case, which corresponds to $z$ being away from the cusps other than $\infty$. Since $z$ is in the Ford domain (3.1), we must have $|bQz - a| \geq 1$ and thus $y \geq \frac{1}{2}Q^{-1}(1 + C_\epsilon Q^\epsilon)^{-2}$. Returning to the first case, if $(c, Q) = 1$ then once again we must have $|Qz - c| \geq 1$ and hence $y \geq \frac{1}{2}Q^{-1}$. Finally, if $(c, Q) > 1$ then we may find natural numbers $k_\pm \leq C_\epsilon Q^\epsilon$ such that $(k_\pm c \pm 1, Q) = 1$. Thus, we have

$$\frac{k_- c - 1}{k_- Q} < x < \frac{k_+ c + 1}{k_+ Q}$$

and the isometric circles $|(k_\pm Q)w + (k_\pm c \pm 1)| = 1$ include the cusp $\frac{c}{Q}$. Hence, we find that every point $z \in \mathcal{F}_{\Gamma, \infty}$ has either $y \geq \frac{1}{2}(1 + C_\epsilon Q^\epsilon)^{-2}Q^{-1}$ or is in a cuspidal region in between to isometric circles of radius at least $C_\epsilon Q^{-1-\epsilon}$. In particular, we find that every isometric circle, which is part of the boundary of $\mathcal{F}_{\Gamma, \infty}$ must have radius at least $\frac{1}{2}(1 + C_\epsilon Q^\epsilon)^{-2}Q^{-1}$. Thus, the side-pairing motions $\left( \begin{smallmatrix} a & b \\ cQ & d \end{smallmatrix} \right)$ of $\mathcal{F}_{\Gamma, \infty}$, which generate $\Gamma$ must have $|cQ| \ll_\epsilon Q^{1+2\epsilon}$, $|a|, |d| \ll_\epsilon Q^{1+2\epsilon}$ and consequently $|b| \ll_\epsilon Q^{1+4\epsilon}$ as $ad - bcQ = 1$. We conclude Theorem 1.

## 4 Graphs

We consider the Brandt–Ihara–Pizer "super singular isogeny graphs", $G(p, \ell)$, where $p, \ell$ are primes with $p \equiv 1 \mod (12)$. They are constructed by interpreting Brandt matrices $B(\ell)$ associated to a maximal order $R$ in the quaternion algebra $B_{p,\infty}$ over $\mathbb{Q}$ ramified at exactly $p, \infty$ as adjacency matrices. They constitute a rich family of non-bipartite $(\ell + 1)$-regular Ramanujan graphs on $n := \frac{p-1}{12} + 1$ vertices [22]. Let $f_j \in L^2(G(p, \ell))$, equipped with probability measure, be an orthonormal eigenbasis of the adjacency matrix $B(\ell)$ with $f_0 \equiv 1$. We may and shall also assume that they are eigenfunctions of all other Brandt matrices $B(m)$ for $(m, p) = 1$. We shall denote the eigenvalue of $f_j$ with respect to $B(m)$ by $\lambda_j(m)$. By identifying the vertices of $G(p, \ell)$ with the class set $B_{p,\infty}^\times \backslash (B_{p,\infty} \otimes \mathbb{A}_f)^\times / (R \otimes \widehat{\mathbb{Z}})^\times$ of $R$, we may interpret the eigenfunctions $f_j$ as automorphic forms in $L^2(\mathrm{PB}_{p,\infty}^\times(\mathbb{Q}) \backslash \mathrm{PB}_{p,\infty}^\times(\mathbb{A}) / K_\infty K_f)$, where $K_\infty$ is a maximal torus and $K_f$ the projective image of $(R \otimes \widehat{\mathbb{Z}})^\times$, constant on each connected component (as a real manifold). The automorphic forms $f_j$ are in one-to-one correspondence with their theta lift, a modular form of weight 2, level $p$, and trivial character, which is cuspidal if and only if $j \neq 0$. They form a basis of Hecke eigenforms of said space. The $m$-th Hecke eigenvalue of the theta lift of $f_j$ is given by the eigenvalue $\lambda_j(m)$ for $(m, p) = 1$ [5, 11]. Thus, by the Petersson trace formula, one has the density estimate (cf. [29, Eq. (4)])

$$\sum_{j \neq 0} |\lambda_j(m)|^2 \ll_\epsilon (p\ell)^\epsilon \left( p + m^{\frac{1}{2}} \right). \tag{4.1}$$

Likewise, the fourth moment bound [15, Theorem 2.1] reads

$$\sup_{x, y \in G(p, \ell)} \sum_{j \neq 0} \left( f_j(x)^2 - f_j(y)^2 \right)^2 \ll_\epsilon p^{1+\epsilon}. \tag{4.2}$$

It is known that the Ramanujan graphs $G(p, \ell)$ have diameter bounded by $(2 + o(1)) \log_\ell(n)$.[4] Here, we shall give an alternative proof which avoids using the Ramanujan bound. Instead, we shall make use of the two inequalities (4.1) and (4.2). For $x, y \in G(p, \ell)$, let $K_t(x, y)$ denote the number of non-backtracking random walks of length $t$ from $x$ to $y$. We have the equality (see [19])

$$\sum_{0 \leq i \leq \frac{t}{2}} K_{t-2i}(x, y) = \frac{1}{n} \sum_j \lambda_j(\ell^t) f_j(x) f_j(y). \tag{4.3}$$

If $x, y$ are of distance larger than $t$, then the left-hand side of (4.3) is zero. We find that

$$\lambda_0(\ell^t) f_0(x) f_0(y) \ll \sum_{j \neq 0} |\lambda_j(\ell^t)| |f_j(x) f_j(y)|, \tag{4.4}$$

---

[4] In fact, sharper results are known, see for example [21].

from which we infer

$$\ell^t \ll \sup_{x \in G(p,\ell)} \sum_{j \neq 0} |\lambda_j(\ell^t)| f_j(x)^2, \tag{4.5}$$

since $\lambda_0(\ell^t) = \frac{\ell^{t+1}-1}{\ell-1}$. By orthonormality, we have $\sum_{x \in G(p,\ell)} f_j(x)^2 = n$. Hence, we may bound the right-hand side further

$$\sup_{x \in G(p,\ell)} \sum_{j \neq 0} |\lambda_j(\ell^t)| f_j(x)^2 \leq \sum_{j \neq 0} |\lambda_j(\ell^t)|$$
$$+ \sup_{x,y \in G(p,\ell)} \sum_{j \neq 0} |\lambda_j(\ell^t)| \left( f_j(x)^2 - f_j(y)^2 \right). \tag{4.6}$$

By applying Cauchy–Schwarz and making use of (4.3) and (4.2), we conclude $l^t \ll_\epsilon n^{2+\epsilon}$ or $t \leq (2 + o(1)) \log_\ell(n)$. In particular, the diameter is bounded by one plus the same quantity, the former getting absorbed into $o(\log_\ell(n))$.

# References

1. Casselman, W.: On some results of Atkin and Lehner. Math. Ann. **201**, 301–314 (1973)
2. Chu, M., Li, H.: Small generators of cocompact arithmetic Fuchsian groups. Proc. Am. Math. Soc. **144**(12), 5121–5127 (2016)
3. Chuman, Y.: Generators and relations of $\Gamma_0(N)$. J. Math. Kyoto Univ. **13**, 381–390 (1973)
4. Eichler, M.: Lectures on Modular Correspondences, vol. 56. Tata Institute of Fundamental Research Bombay, Mumbai (1955)
5. Eichler, M.: The basis problem for modular forms and the traces of the Hecke operators. In: Modular Functions of One Variable, I (Proceedings of International Summer School, University of Antwerp, Antwerp, 1972). Lecture Notes in Mathematics, vol. 320, pp. 75–151. Springer, Berlin (1973)
6. Ford, L.R.: The fundamental region for a Fuchsian group. Bull. Am. Math. Soc. **31**(9–10), 531–539 (1925)

7. Frączyk, M., Raimbault, J.: Betti numbers of Shimura curves and arithmetic three-orbifolds. Algebra Number Theory **13**(10), 2359–2382 (2019)

8. Frasch, H.: Die Erzeugenden der Hauptkongruenzgruppen für Primzahlstufen. Math. Ann. **108**(1), 229–252 (1933)

9. Golubev, K., Kamber, A.: Cutoff on graphs and the Sarnak–Xue density of eigenvalues. Eur. J. Comb. (2022). https://doi.org/10.1016/j.ejc.2022.103530

10. Golubev, K., Kamber, A.: Cutoff on hyperbolic surfaces. Geom. Dedicata **203**, 225–255 (2019)

11. Hijikata, H., Pizer, A.K., Shemanske, T.R.: The basis problem for modular forms on $\Gamma_0(N)$. Mem. Am. Math. Soc. **82**(418), 1–159 (1989)

12. Iwaniec, H.: Spectral Methods of Automorphic Forms. Graduate Studies in Mathematics, vol. 53. American Mathematical Society, Providence; Revista Matemática Iberoamericana, Madrid, 2nd edn (2002)

13. Jacquet, H., Langlands, R.P.: Automorphic Forms on GL(2). Lecture Notes in Mathematics, vol. 114. Springer, Berlin (1970)

14. Johansson, S.: On fundamental domains of arithmetic Fuchsian groups. Math. Comput. **69**(229), 339–349 (2000)

15. Khayutin, I., Nelson, P.D., Steiner, R.S.: Theta functions, fourth moments of eigenforms, and the sup-norm problem II. Arxiv Preprint (2022). arXiv:2207.12351

16. Khoai, H.H.: Sur les séries $L$ associées aux formes modularies. Bull. Soc. Math. Fr. **120**(1), 1–13 (1992)

17. Kulkarni, R.S.: An arithmetic-geometric method in the study of the subgroups of the modular group. Am. J. Math. **113**(6), 1053–1133 (1991)

18. Kurth, C.A., Long, L.: Computations with finite index subgroups of $PSL_2(\mathbb{Z})$ using Farey symbols. In: Advances in Algebra and Combinatorics, pp. 225–242. World Scientific Publishing, Hackensack (2008)

19. Lubotzky, A., Phillips, R., Sarnak, P.: Hecke operators and distributing points on $S^2$. II. Commun. Pure Appl. Math. **40**(4), 401–420 (1987)

20. Lubotzky, A., Phillips, R., Sarnak, P.: Ramanujan graphs. Combinatorica **8**(3), 261–277 (1988)

21. Nestoridi, E., Sarnak, P.: Bounded cutoff window for the non-backtracking random walk on Ramanujan graphs. Arxiv Preprint (2021). arXiv:2103.15176

22. Pizer, A.K.: Ramanujan graphs. In: Computational Perspectives on Number Theory (Chicago, IL, 1995). AMS/IP Studies in Advanced Mathematics, vol. 7, pp. 159–178. American Mathematical Society, Providence (1998)

23. Rademacher, H.: Über die Erzeugenden von Kongruenzuntergruppen der Modulgruppe. Abh. Math. Sem. Univ. Hamburg **7**(1), 134–148 (1929)

24. Ratner, M.: The rate of mixing for geodesic and horocycle flows. Ergod. Theory Dyn. Syst. **7**(2), 267–288 (1987)

25. Reidemeister, K.: Knoten und Gruppen. Abh. Math. Sem. Univ. Hamburg **5**(1), 7–23 (1927)

26. Rickards, J.: Improved computation of fundamental domains for arithmetic Fuchsian groups. Arxiv Preprint (2021). arXiv:2110.11503

27. Saha, A.: Sup-norms of eigenfunctions in the level aspect for compact arithmetic surfaces. Math. Ann. **376**, 609–644 (2019)

28. Sarnak, P.: Letter to Scott Aaronson and Andy Pollington on the Solovay–Kitaev theorem. Institute for Advanced Study (2015). http://publications.ias.edu/sarnak/paper/2637

29. Sarnak, P., Zubrilina, N.: Convergence to the Plancherel measure of Hecke Eigenvalues. Arxiv Preprint (2022). arXiv:2201.03523

30. Schreier, O.: Die Untergruppen der freien Gruppen. Abh. Math. Sem. Univ. Hamburg **5**(1), 161–183 (1927)

31. Selberg, A.: On the estimation of Fourier coefficients of modular forms. In: Proceedings of Symposia in Pure Mathematics, vol. VIII, pp. 1–15. American Mathematical Society, Providence (1965)

32. Voight, J.: Computing fundamental domains for Fuchsian groups. J. Théor. Nombres Bordeaux **21**(2), 469–491 (2009)